

Web Services en el contexto de la Visualización de Grandes Volúmenes de Datos

Mauricio Cele López

Silvia Mabel Castro

Sergio Rubén Martig

VyGLab - Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica

Dpto. de Ciencias e Ingeniería de la Computación

Universidad Nacional del Sur

Av. Alem 1253, (8000) Bahía Blanca, Argentina

RESUMEN

Gracias a los crecientes avances tecnológicos se aprecia un crecimiento exponencial en densidad y tamaño de los volúmenes de datos; la visualización de tales cantidades de datos puede sobrepasar rápidamente el poder de procesamiento local. Con la disponibilidad de la GRID, la ubicuidad de Internet y el soporte de los Web Services se han propuesto varios sistemas de visualización distribuida que asisten a científicos y usuarios de todo el mundo. Sin embargo, algunos de estos sistemas se han planteado desde el punto de vista de la colaboración a distancia y otros desde el punto de vista de compartir recursos; debido a esto, los sistemas existentes no están preparados para manejar volúmenes de datos realmente grandes. Para contar con un sistema de Visualización en la Web es imperativo hacer uso eficiente de la red; la estrategia consiste en transmitir solo las porciones de datos que serán procesadas y visualizadas; esto implica crear una representación de datos adecuada que permita extraer datos con diversos niveles de detalle para poder transmitirlos y visualizarlos progresivamente. Finalmente esto debe integrarse en un sistema de caché distribuido que evite retransmisiones innecesarias y haga que los datos residan más tiempo donde más se necesitan.

Keywords: Web Services, Visualización Distribuida, Visualización Científica, Representaciones de Datos Volumétricos, Rendering de Volúmenes.

INTRODUCCION

Actualmente la comunidad científica necesita visualizar enormes cantidades de datos volumétricos, muchas veces multidimensionales, multivaluados y variantes en el tiempo. El proceso de descubrimiento de características en tales datos es inherentemente iterativo y se realimenta constantemente para mejorar cualitativa o cuantitativamente la vista, aumentando la semántica. Es deseable que el tiempo de respuesta del sistema esté en relación directa con la calidad de la

visualización obtenida, dando libertad al científico de escoger la calidad deseada en ciertas etapas del análisis donde necesita tiempos de respuesta muy bajos.

Esta es una tarea compleja y computacionalmente intensiva que, normalmente, se puede dividir en subprocesos con el objetivo de estructurar y distribuir la carga de trabajo tanto conceptual como físicamente. Se han propuesto modelos conceptuales del proceso de visualización (ver [5]) que describen de forma general los datos, sus transformaciones, estados y las interacciones del usuario; estos modelos son adecuados para modelar sistemas de visualización con un alto nivel de abstracción. Por otro lado, el desarrollo actual de los modelos físicos del proceso de visualización distribuida recaen en Web Services que proveen interoperabilidad entre distintas plataformas y lenguajes de programación; en representaciones de datos adecuadas para tratar con grandes cantidades de datos en múltiples resoluciones o niveles de detalle y en la eficiencia y confiabilidad de los métodos de representación, por nombrar algunos. Todo esto plantea un gran reto para la comunidad de visualización que, si bien ha encontrado soluciones parciales, aun tiene mucho camino por recorrer.

A pesar del incremento en la velocidad de las redes, el mayor inconveniente en la visualización distribuida reside en la ineficiencia de transmisión de datos por la red, de modo que, para que sea viable, es necesario desarrollar nuevas estrategias para minimizar el tráfico de datos, encontrar representaciones de datos adecuadas para su transmisión progresiva así como algoritmos de rendering preparados para trabajar sobre esas representaciones. Nuestra línea de investigación se centra en esos tres aspectos.

TRABAJO RELACIONADO

McCormick, de Fanti y Brown [6] sientan las bases de la visualización distribuida en 1987. En su informe plantean el uso de una red de servidores para hacer cálculos científicos mientras el rendering

se hacía en las estaciones cliente. McCormick *et al* plantea los *Ambientes de Visualización Modular* (MVE), los cuales son módulos separados que realizan tareas como leer datos, calcular iso-superficie o renderizar. Este modelo incluye un sofisticado editor para construir aplicaciones, en el cual se escoge visualmente qué módulos se usarán y cómo se enlazarán, es decir, el usuario crea su propia *línea de producción* o pipe line. Ejemplos de estas aplicaciones son IRIS Explorer e IBM Data Explorer. Los módulos pueden ser implementados como procesos separados, es decir, pueden ejecutarse en paralelo. Si es así, fácilmente se adecuan a un modelo de sistema distribuido. La Figura 1 ilustra el enlace de tres módulos para la visualización de iso-superficies poligonales:

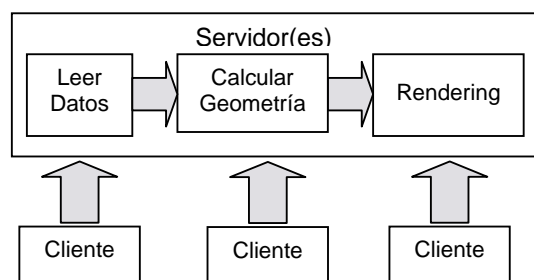


Figura 1

Ang *et al.* [1] en 1994 propusieron un sistema de visualización basado en la Web. Los datos residen en los servidores y son transmitidos a las máquinas cliente como attachments MIME. La visualización se hace usando el poder de procesamiento de la máquina cliente. Posteriormente Yeo, en 1998 [7] extendió el trabajo de Ang *et al* dándole más énfasis al procesamiento en los servidores; la geometría generada remotamente era transmitida al cliente como attachments MIME y el rendering se hacía localmente.

Más recientemente, en el 2000, Engel [4] plantea un sistema de visualización dinámico basado en Applets de Java. El usuario descarga de Internet la aplicación cliente, transmite los datos al servidor vía FTP, éste procesa los datos y retorna un archivo VRML para que el rendering se realice localmente (browser). En este trabajo ya se reporta la idea de transmitir progresivamente niveles de detalle al cliente con el objetivo de disminuir el tiempo de respuesta.

El servicio web de Visible Human Project ofrecido por EPFL ofrece un Applet como interfaz para seleccionar slices de interés del dataset; el pedido es procesado por los servidores, los cuales extraen las porciones y las retornan al Applet para su visualización local.

Tanto Engel como Ang y Yeo asumen que el cliente tiene recursos y poder de procesamiento para

realizar el rendering. Esto es esencial ya que sus sistemas proporcionan visualización en tiempos interactivos, tal que el usuario puede ver reflejadas inmediatamente sus interacciones. Sin embargo los clientes que no tienen poder de procesamiento no están contemplados.

Otros investigadores proponen que el proceso de visualización se haga del lado del servidor pero dirigido por el cliente; bajo este paradigma los datos residen y se procesan en los servidores y el resultado es transmitido al cliente para un mínimo procesamiento final. Un ejemplo es el sistema CurVis [3] que produce visualizaciones de *Dinámica de Fluidos 2D* del lado del servidor y transmite el resultado como una imagen comprimida a un cliente móvil, con lo cual, lógicamente, el cliente pierde interactividad.

LINEAS DE INVESTIGACION

Existen muchos problemas no resueltos en la visualización directa de volúmenes, uno de ellos es la visualización interactiva de enormes volúmenes de datos. Si bien este problema no tiene una solución a corto plazo, nuestro objetivo principal apunta a dar un paso más en el camino hacia la interactividad, esperando que en un futuro el aumento en la velocidad de las redes habilite tiempos cada vez más cortos de transmisión.

Teniendo en cuenta el creciente tamaño de datos a visualizar es imperativo hacer uso eficiente de la red. Para ello nuestra estrategia consiste en encontrar una representación de datos adecuada que encapsule varios niveles de detalle tal que, cada nivel de detalle se pueda transmitir por separado; transmitiendo primero los datos con menor nivel de detalle se desea reducir el tiempo que espera el usuario antes de apreciar una imagen y poder manipularla. Al transmitir los subsiguientes niveles de detalle se espera aumentar la calidad de la visualización.

Esta forma de visualización progresiva tiene una ventaja fundamental sobre otras aproximaciones: no se retransmitiría la misma información más de una vez. Simplemente cada nivel de detalle añadiría nueva información a la anterior.

De este modo se lograría una verdadera distribución de datos a través de la red, con menor cantidad de datos repetida en cada servidor.

La visualización progresiva de datos es también un tema de investigación. Es necesario desarrollar algoritmos de rendering directo de volúmenes adecuados para mostrar datos en su nueva representación y que, además, puedan soportar su enorme tamaño. Se deberá adecuar o bien desarrollar

algoritmos de memoria externa para todos los procesos que implica el Rendering.

Para disminuir aún más el tráfico de la red el sistema debe implementar un sistema de caché de alto nivel que mantenga copias limpias de los datos en los Servidores donde tienen más demanda. La demanda de datos varía según las tareas repetitivas que realiza el usuario.

Actualmente se está estudiando el uso de Wavelets como representación de volúmenes de datos que encapsula varios niveles de detalle (ver [2]) así como el uso de algoritmos proyectivos de rendering de volúmenes Out-Of-Core para su visualización.

Todo el proyecto se enmarca en el Modelo Unificado de Visualización (MUV) de Martig *et al* [5], que nos permitirá sentar las bases del sistema con un alto nivel de abstracción y sin perder generalidad. Esto es muy importante para la futura integración del proyecto a otros desarrollados en este mismo marco conceptual y a su vez aportar información que ayude a completar el MUV.

AGRADECIMIENTOS

El presente trabajo se está llevando a cabo en el Laboratorio de Investigación y Desarrollo en Visualización y Gráfica, Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur y es parcialmente soportado por el proyecto de Investigación 24/N015, *Modelo Unificado de Visualización. Operadores y Operandos*, otorgado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional del Sur.

BIBLIOGRAFÍA

- [1] Ang C. S., Martin D. C., Doyle M. D., “*Integrated Control of Distributed Volume Visualization Through the World Wide Web*”. In Proceedings of IEEE Visualization '94 (1994), IEEE Computer Society Press. 22.
- [2] Castro S., Castro L., Boscardín L., De Giusti A., “*Multiresolution Wavelet Based Model for Large Irregular Volume Data Sets*”, in Proceedings of WSCG'2006, pp 101–108, 8, 2006.
- [3] CurVis Web Site.
<http://www.informatik.uni-rostock.de/Projekte/movi/proto.html#CURVIS>
- [4] Engel K., Ertl T., “*Texture-based Volume Visualization for Multiple Users on the World Wide Web*”. Proceedings of the Eurographics Workshop in Vienna, Austria (1999), pp. 115–124. 21.
- [5] Martig S, Castro S, Fillottrani P, Estévez E., “*Un Modelo Unificado de Visualización*”, Proceedings, pp. 881-892, 9° Congreso Argentino de Ciencias de la Computación. 6 al 10 de Octubre de 2003. La Plata. Argentina.
- [6] McCormick B. H., Defanti T. A., Brown M. D., “*Visualization in Scientific Computing*”. Computer Graphics 21, 6 (1987). 2.
- [7] Yeo A., “*Client-based Web Visualization*”. Master's thesis, School of Computer Studies, University of Leeds, U.K., 1998. 7, 22.